Trois livres sur l'intelligence artificielle — IA faible, IA forte

mardi 21 janvier 2020, par KRIVINE Hubert (Date de rédaction antérieure : 17 janvier 2020).

L'intelligence artificielle (IA) fait de plus en plus fantasmer, promettant le pire (un monde où les machines domineraient les humains) comme le meilleur (un monde enfin débarrassé de toutes les tâches peu gratifiantes). Un cadrage à partir de trois livres :

• *Quand la machine apprend, La révolution des neurones artificiels et de l'apprentissage profond*, Yann Le Cun, Odile Jacob, 2019.

• *Intelligence artificielle, vers une domination programmée ?*, Jean-Gabriel Ganascia, Le cavalier Bleu, 1917.

• *L'intelligence artificielle n'existe pas*, Luc Julia, First éditions,1919.

Sommaire

- IA faible, IA forte
- De la corrélation à la causali
- En conclusion

Ces trois livres ont en commun d'avoir été écrits par des auteurs qui sont parmi les meilleurs spécialistes français de l'intelligence artificielle. Ils ont aussi en commun un mérite essentiel : celui de doucher les ardeurs de ceux qui extrapolent sans limite les succès actuels de l'IA. Citons parmi les plus connus Elon Musk, Bill Gates, voire Stephen Hawking.

IA faible, IA forte

L'usage est de distinguer l'IA faible – ou mieux dit, étroite – de l'IA forte. La première doit résoudre un problème ou prendre une décision dans un cadre bien défini (jeu de go, conduite automatique de trains, reconnaissance d'image, traduction…), c'est-à-dire avec peu d'interaction avec l'environnement. C'est la seule réellement existante aujourd'hui. La seconde viserait à émuler le cerveau humain. C'est elle qui assurerait la fameuse "singularité" c'est-à-dire le triomphe de la machine sur l'Homme. Par exemple, l'IA faible vous indiquera le meilleur itinéraire pour aller de chez vous à un point B. L'IA forte vous dirait, tenu compte de votre situation géographique, financière, psychologique, de la météo et de l'opinion de votre conjoint, quel point B choisir. Notons que la publicité ciblée donne déjà une image très embryonnaire de ce pouvoir de l'IA.

Les trois auteurs considèrent évidemment à juste titre que l'IA faible est bien supérieure à l'humain. Ses succès au jeu d'échec et au jeu de go sont spectaculaires. C'est par exemple en médecine la caractérisation de tumeurs cutanées ou le diagnostic de la rétinopathie diabétique meilleure que celle des médecins. Sans parler de la reconnaissance faciale si prisée des policiers de tous pays. Mais pour ces trois auteurs, n'importe quelle souris est infiniment plus intelligente, parce qu'elle

saura s'adapter à bien des situations imprévues. Ce qui est difficile, voire impossible, pour l'IA puisqu'elle procède par induction, c'est-à-dire par généralisation de situations répertoriées.

L'IA moderne, quand elle est fondée sur l'apprentissage machine (*machine learning*), se nourrit de données massives (*big data*), c'est-à-dire qu'elle effectue un traitement statistique de millions (voire de milliards) de données brutes, étiquetées ou pas, permettant de révéler des corrélations insoupçonnées. Il convient bien sûr que les données ingurgitées par la machine ne soient pas biaisées ou incomplètes pour une raison ou une autre.

De la corrélation à la causalité

L'IA permet de prévoir, pas de comprendre : la corrélation peut signer une causalité, elle ne l'implique pas. Le passage de l'une à l'autre présuppose de grandes précautions méthodologiques et le passage de la causalité à l'explication (que ne semble pas considérer nos auteurs) impose une réflexion théorique. Il y a plus : seule cette dernière peut permettre des prévisions inouïes, au sens étymologique (qu'on n'a jamais ouï, entendu). En effet, pour l'IA qui fonctionne par généralisation, tout ce qui se produira s'est soit déjà produit, soit est proche de l'avoir été. Jamais, n'on aurait pu, par exemple, anticiper l'existence des ondes radio ou des ondes gravitationnelles. Gageons que comme en 2008, les prévisions de l'IA pour la crise économique à venir dont les préconditions sont multiples et pour certaines originales seront fausses.

En tirant un peu les considérations de ces trois auteurs, il me semble qu'on peut conclure que l'IA est d'autant plus efficace que les problèmes qu'elle envisage sont simples. Les problèmes que nous baptiserons ici de « simples » sont ceux dont l'étude peut raisonnablement être isolée de leur environnement. C'est typiquement le cas des jeux (go, échecs…) ou de la reconnaissance d'images. Simple ici n'a pas du tout la signification habituelle de aisé ou de facile : il y aurait sensiblement 10600 parties [1] possibles au premier jeu et « seulement » 10123 au second ! Notons que le nombre total d'atomes de l'univers est estimé à 1080. Simple veut dire sans intervention inopinée ou compliquée de l'extérieur. C'est souvent le cas en physique où on peut isoler son objet d'étude de l'environnement (voir mon article dans précédent à ce sujet), mais allez le faire en économie, en histoire, en pédagogie [2], en biologie [3] ou en théorie de l'évolution !

Bien qu'éventuellement non causales, ces corrélation peuvent avoir des vertus prédictives. Un exemple : plus les gens consomment de glace, plus il y a de noyades. Pourquoi ? Parce que la consommation de glace est liée à la température et plus il fait chaud, plus les gens se baignent…

Yan Le Cun, à juste titre, souligne que le passage à la causalité est un des principaux problèmes posé à l'IA, mais il n'en dit pas beaucoup plus. Les médecins ont ce problème : aspirine et baisse de fièvre sont corrélées. Ici la causalité peut s'établir (assez) facilement : on divise les malades en deux lots. Au premier on donne de la poudre de perlimpinpin (par exemple, une préparation homéopathique) et au second l'acide salicylique. C'est ce qu'on appelle un raisonnement contra-factuel. Mais il est très souvent inapplicable : on n'est pas maître de trier ses données.

Dans les domaines "simples", les prévisions ou les indications de l'IA, bien que déterminées de façon statistique, peuvent être pertinentes individuellement. Aux échecs, la partie sera – presque – à coup sûr gagnée par la machine. Par contre, ailleurs, dans les domaines en interaction forte avec l'extérieur, les prévisions sont pertinentes seulement en moyenne. Il y a là une porte ouverte à l'imprévu et à l'originalité que l'IA ne domine pas.

Certes on peut supposer que les mêmes causes produisent les mêmes effets, mais tout événement est singulier. Où et quand y a-t-il strictement les mêmes causes ? Il peut se faire que des variations

indécelables de causes peuvent générer des variations colossales des effets. Par exemple, les résultats d'un pile ou face sont très fortement dépendants de l'environnement (les conditions du lancer, l'agitation de l'air, etc.). L'IA vous indiquera qu'en moyenne 50 % des lancers donneront pile, mais rien sur un lancer particulier. De même en économie, l'homme "moyen" n'existe pas. Les probabilités de récidive d'un détenu n'impliquent pas non plus d'anticiper le comportement d'un futur libéré. L'homme réel a toujours une distance avec l'homme moyen et cette distance est déterminée par son interaction avec l'environnement qui est toujours singulière.

Pourquoi l'IA n'est pas intelligente

Admettons la causalité établie : l'aspirine est bien la cause de la chute de la température. Reste maintenant à expliquer le mécanisme de cette action. Bien entendu, une explication est toujours relative : on peut dire que la loi de la gravitation de Newton est cause de la chute des pommes comme du mouvement des planètes. Mais qu'est-ce qui cause cette loi ? Einstein expliquera que c'est une courbure de l'espace-temps créée par les masses en présence. Bon, on le voit, pour les enfants comme pour les grands, la suite des pourquoi semble infinie.

La machine est dédiée à une tâche (éventuellement très complexe), le cerveau ne l'est pas. Il est le produit d'une évolution darwinienne sur des centaines de millions d'années qui lui ont « appris » à faire face à toutes sortes de situations. Loin d'être linéaire, cette évolution a exploré bien des chemins de traverse, apparemment inutiles, mais qui lui ont conféré par un processus encore bien mystérieux son imagination et sa créativité. Par exemple pour reconnaître un hippopotame, une machine a besoin de s'être entraînée sur des milliers d'images, il en faudra deux ou trois pour un enfant de 4 ans. Voilà une vraie interrogation dont les trois auteurs sont conscients. Clairement pour Yann Lecun, "l'inné" est inévitable. Il serait donc une propriété émergente, acquise sur 3 milliards d'années. Voilà un vaste sujet d'études.

L'IA forte, si jamais elle peut se concrétiser, reste loin, très loin, d'être à un horizon prévisible. Ceux qui y croient néanmoins commettent, à mon avis, la double erreur :

i) de la réduire à une agrégation d'IA faibles,

ii) d'identifier la connaissance de l'environnement au contenu des data, si big soient-elles.

En conclusion

Outre son titre justement provocateur, et ses considérations raisonnables sur l'IA, le livre de Luc Julia a ceci de particulier qui énervera probablement les lecteurs de cette revue lorsqu'il juge que « Les capacités d'écoute et d'analyse du président Macron sont impressionnantes », à tel point qu'il aura su séduire Sohn, le président de Samsung. On apprend que pendant les cinq premières années de la présidence de François Hollande des réunions de grosses compagnies étrangères étaient organisées par Muriel Pénicaut, présidente de Business France et cordonnées à l'Élysée par Macron. Et que ça continue de plus bel depuis.

Le livre de J-G Ganascia est peut-être le plus convaincant quand il s'attaque au mythe de la singularité, celui qui nous décrit un monde à venir dominé par des machines au détriment des humains. Il ouvre par ailleurs de bonnes question théoriques, notamment sur la créativité potentielle de la machine.

Ce qui fait – entre autres – la force du livre de Le Cun est son exposé très pédagogique des mécanismes du deep learning (l'apprentissage profond). C'est un livre riche à tout point de vue.

Hubert Krivine

P.-S.

• Créé le Vendredi 17 janvier 2020, mise à jour Vendredi 17 janvier 2020, 12:20 : https://npa2009.org/idees/culture/trois-livres-sur-lintelligence-artificielle

• Hubert Krivine est physicien, dernier ouvrage paru : *Comprendre sans prévoir, prévoir sans comprendre*, Cassini, 2018.

Notes

[1] C'est-à-dire 1 suivi de six cents zéros !

[2] Voir les déboires de la neuropédagogie, quand elle prétend ramener le fonctionnement du cerveau à sa physiologie.

[3] Ecouter la très belle interview de Michel Blay publiée dans Mediapart du 27 août 2019 : https://www.mediapart.fr/journal/culture-idees/270819/derriere-les-neuropedagogies-une-ideologie-individualiste